DOCUMENT RESUME

ED 166 222 TM 008 170

AUTHOR Kennedy, Mary M.
TITLE Generalization of Findings From Single Case Studies.
PUB DATE Mar 78
NOTE 26p.; Paper presented at the Annual Meeting of the American Educational Research Association (62nd, Toronto, Ontario, Canada, March 27-31, 1978); Not available in hard copy due to marginal legibility of original document

EDRS PRICE MF-$0.83 Plus Postage. HC Not Available from EDRS.
DESCRIPTORS *Case Studies; *Data Analysis; Educational Research; *Evaluation Criteria; *Evaluation Methods; *Generalization; Guidelines; *Research Methodology; Sampling; Statistical Analysis
IDENTIFIERS *Sample Size

ABSTRACT

Although single case studies might be useful to evaluators for a variety of purposes, there are no generally accepted ways for drawing inferences about the generality of findings from a case study. Single case studies are defined in this paper as either studies of single events, or disaggregated studies of multiple events. The data may be qualitative or quantitative, and may be derived from controlled experiments or from observation. There are two spans to the bridge of inference. The statistical span connects the experimental sample to a popluation just like that sample. The second span connects the population to a group judged to be sufficiently similar. In case law or in clinical practice, the judgment of sufficient similarity --that is, the judgment of the appropriateness of the generalization--is made by the user. This application of single case data may also be appropriate in educational evaluation. (CTM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

- ED166222

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Generalization of Findings From Single Case Studies

Mary M. Kennedy

U.S. Office of Education

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY
Mary Kennedy
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

(Printed in the U.S.A.)

Paper presented to the American Educational Research Association, March 1978, Toronto, Ontario.

The opinions expressed in this paper are those of the authors and should not be construed to represent those of the U.S. Office of Education.

TM008 170

Generalization of Findings from Single-Case Studies

Mary M. Kennedy

I chose the topic of generalizability because it seemed to me that, although single case studies might be useful for a variety of purposes, they presented difficulties to any evaluator who wished to generalize his findings. There are no generally accepted ways for drawing inferences about the generality of findings from a case study, or even from studies of a very few cases. Advocates of single case methodologies have used a variety of arguments to overcome the problems of sampling limitations, but none has satisfied those who rely on multiple cases to draw generalizable findings from studies. Some authors (e.g., Feinberg, 1977) have erroneously assumed that by increasing the number of data points on a single case, one has eliminated the problem. But this is not a solution, for these several data points are still based on only one subject. Other authors have argued the importance of single-case methodologies because they can accomodate alternativeepistemological viewpoints (e.g., Rist, 1977; Stake, 1978). In these discussions, arguments are made for the validity of qualitative data, subjective impressions, or descriptions of naturally occuring events. While these arguments may be valid, they suggest that the evaluator needs the talent of Tolstoi to be able to describe these events in ways that allowed a reader to draw the appropriate inference. Still other authors (e.g., Edgar and Billingsley, 1974; Edgington, 1967) have argued for the application of nonstatistical, but still logical, rules of generalization. It is important to realize that non-statistical arguments need not be invalid. Yet, many researchers may be timid about attempting such inferences simple because the rules as to what constitutes a reasonably sound inference are ambigous, relative to the rules as to what constitites a sound statis-

tical inference. What is needed are rules of inference that reasonable people can agree on. If these rules were applied and revised over time, it might be possible for us to be as comfortable with such an inferential process as we are with the statistical inferential process. In this paper, I will address myself to such a set of rules, and will focus only on rules that apply when samples are restricted. By doing so, I will not concern myself with a number of other issues that can be raised regarding single-case methodologies, such as the rules by which causal inferences are drawn, or the way in which data are collected. In fact, there are a number of issues that are important, but that I will not get into, so it might be well for me to start by listing some of these.

First, I won't talk about quantitative versus qualitative observations. Either kind of observations can be made in single case or multiple case studies. Though the issue is an important one, it is not necessarily related to the size of the sample.

Second, I will not address the relative values of controlled experiments versus observation of naturally occurring events. Again, either approach can be used on any sample size. I will, however, assume in this discussion that there is a treatment of interest, and that the goal of the study is to generate some form of generalizable knowledge about that treatment.

Finally, I hope not to go into the problem of the appropriateness of different units of analysis, except to the extent that the choice of units may prove to be related to generalizability. It is just as possbile to select a wrong unit in a single-case study as in a multiple-case study, although it may be easier in a single-case study to see the fallacy of wrong reasoning when it is time to generalize about treatment effects if the wrong unit is used.

There is one topic related to the single versus multiple case study, however, that I cannot ignore, and that is the subject of aggregation. For purposes of this paper, I will define single case studies as either (a) studies of single events, or (b) disaggregated studies of multiple events. That is, it is possible to study more than one case, but to study them individually, rather than averaging or in other ways pooling the data across cases. The reason that I make this distinction is that the disaggregated study of multiple cases is not often considered as an alternative, and it may prove to be a relatively strong approach to evaluation. I will consider the problems of generalization from either of these two types of single-case studies.

The reason it is necessary to specify that these several topics will not be discussed is that they often are discussed in debates about the merits of single-case studies. It seems necessary in these debates to describe the attributes that these studies could have that are not attributes of some other types of studies. For my purposes, I would like to confine myself solely to differences of sample size. Thus, the comparison against which I will contrast single studies can simply be labled the "multiple-case study". Multiple-case studies are popular because they allow the application of a variety of statistical techniques to the data to determine the generalizability of the findings. With proper sampling techniques, the variance of the sample can be used to estimate the variance of the population, and it has even been argued (e.g., McNemar, 1940) that the existence of human variation forces psychological research to rely on sampling. Clearly a single-

case study cannot meet this criterion. And without an estimate of population variability, no basis exists for statistical inference about the population (Edgington, 1967).

This single weakness of single-case studies is sufficient to discourage many researchers from employing them, in spite of demonstrations of the influence single-case studies have had on theory (Dukes, 1965), and a variety of other arguments as to their advantages (Herson and Barlow, 1972; Edgar and Billingsley, 1974).

In this paper, I will attempt to explicate what some alternate rules for generalization inferences might be. The rules that I will offer are far from being complete or fully developed, but may offer a base which can later be refined by practicing researchers.

The Problem of Generalization

Let me start by reminding you that an inference of generalization is always tentative. That is, data might offer confirming on disconfirming evidence, but never conclusive evidence. Not even in multiple-case studies can the evaluator generate conclusive evidence of generalizability. The strength of the evidence is a matter of judgment. For that reason, a good evaluator tries to make clear the strength of his evidence. A currently popular term for the generalizability of a finding is external validity (Campbell and Stanley, 1963), but a more appropriate term would be "strength of external validity", or "strength of generalizability", since these terms suggest that generalization is a judgment of degree, rather than by a binary decision.

A second point about generalization is that it is not simply a function of the number of units one has observed. More important are the kinds of units observed, that is, the range of characteristics of the units investigated and the range of conditions under which observation occurred. The range of characteristics included in a sample increases the range of population characteristics to which generalization is possible. Thus, generalizations may vary in their range as well as in the strength of their arguments. It should be clear, then, that a wider range of generalization is not necessarily achieved by increasing the sample size. Large samples may be selected for their geographic convenience, their political accessibility, or other irrelevant factors, and may consist of highly homogeneous groups. For example, an investigation of treatment effect on 100 students in a "college town" elementary school may have a narrower range of generalization than a study of 10 children whose parents range in income and education levels and whose homes are in rural, suburban, and urban areas. Cornfield and Tukey (1956) described inference as having two spans. One, a statistical span, connects the sample to a population just like the sample. The second span connects to a population believed or assumed to be sufficiently similar that the study findings apply there as well. The second span cannot rely on statistics for assistance, but is not necessarily less valid. I think when we consider single case studies, we are forced to rely almost completely on the second span for our inferences. And the rules for generalizing across that span are much more ambiguous than are the statistical rules for generalizing across the first span. I am not capable of providing such a set of rules, but

would like to take this opportunity to offer a starting point for generalization in each of our two kinds of single-case studies, beginning with designs with disaggregated replications of single-cases.

Generalizing from Replicated Single-Case Studies

I begin with the problem of generalizing from replicated single cases, primarily because these designs are easier to generalize from than are designs with single case alone. For purposes of this discussion, however, let us assume that the number of replications is still quite small -- somewhere between 2 and 10, and consider some criteria that might be used to generalize from these cases. Criteria for generalizing begin with the attributes of the sample cases.

Criteria for Sample Attributes

1. Wide Range of Attributes across the Sample Cases.

I already referred to this criterion earlier and mentioned that it applied to group studies as well as single-case studies. What is important here, though, is to recognize that it is possible to encompass a wide range of attributes with only a few cases. Six school districts can vary considerably in their size, income, and in their community's educational levels. Six children can similarly vary. But because of the possibility in these designs of confusing idiosyncratic outcomes with generalizable outcomes, we need other criteria to strengthen the generalization.

2. Many Common Attributes between Sample Cases(s) and the Population of Interest

To the extent that we can identify all those attributes that are

common between the sample and one or more populations of interest, these common features may form a basis for generalization. To determine the common attributes, the evaluator may need to define the attributes of population(s) of interest. This can be done in one of three ways -- the evaluator may rely on prior knowledge of the population attributes, he may make assumptions about its attributes, or he may be able to define a hypothetical population of interest by its attributes. Any of these three approaches is appropriate, and all are probably equally feasible. I say this on the assumption that no evaluation is done in a vacuum, but instead follows on a continuing stream of studies, each of which has contributed in some way to the body of knowledge that the evaluator of a new project has at his disposal.

Now there are obviously situations where an evaluator may want to generalize but not to any particular population. That is, he merely wants to have his findings apply as much as possible to other situations. It would still be possible for such an evaluator to describe several "common" features of his sample. For example, he can identify all of the normative characteristics of his sample cases - a child's IQ or reading level, a school district's wealth or size, and argue that his findings generalize to "similar" cases. Now, the first criterion, the range of attributes in the sample cases, increase the range of generalization. This second criterion, the number of attributes in common with the population, increases the strength of the generalization. But, of course, not any attributes will do. If irrelevant attributes were employed, and dozens listed off, they would do

little to strengthen our case. I will come to a criterion of relevance later on.

3. Few Unique Attributes in the Sample Case(s)

This criterion is actually the converse of the preceding. To the extent that our sample has unique attributes, these attributes interfere with generalizability. But unique attributes are more difficult to isolate. They might include special circumstances, for example, such as the recent death of a child's parents, or an unusual method by which school board members are selected in a sample district. Unique attributes may often be attributes that one would never think to look for. Yet one must attempt to find these, for it is only be separating the unique from the common features that the relationship between a sample and a population can be defined. It would be acceptable, I think, to define unique attributes post hoc. At least it would be preferable to not defining unique attributes at all.

4. Relevance of Attributes

We have considered three criteria for sample attributes that are necessary for generalizing. The first was the range of attributes included in the sample, the second was the number of similarities between the sample and the population(s) of interest. Our third was the lack of unique features. This fourth now follows on the first three and requires that the attribute previously identified be relevant.

In educational studies, for example, attributes such as children's hair color would not be relevant. Neither would attributes of schools such as the brand of their furnaces. Determining which attributes are

relevant, then, requires some analysis of the treatment itself, or of the hypotheses or study questions of interest. For starters, one would want to list:

(a) those attributes which the treatment is designed to influence (e.g., a child's reading level, or a school's compliance with a federal regulation)

(b) those attributes known from prior experience to be related to the first set of attributes (e.g., the educational level of the child's parents; the resources available to the school)

(c) those attributes hypothesed by other researchers or evaluators to be related, but which lack substantial evidence of relationship.

Where sufficient prior knowledge is available, these attributes can also be characterized by the strength of the evidence that they are causally related, and by the nature of their apparent relationship to the outcome of interest (e.g., necessary but not sufficient condition, neither necessary nor sufficient, etc.)

Notice that these criteria for revelance could only be useful for common attributes, and not for unique. The relevance of unique attributes cannot be determined on the basis of prior experience, precisely because the feature is unique. The relevance of unique attributes can only be determined by an analysis of their relationship to the treatment during the course of the study. To see how this might be done, let us now turn from arguments of generalization that are based on the sample to those that are based on the treatment.

Criteria for Attributes of the Treatment

Characteristics of the treatment are rarely considered as useful evidence for generalization, perhaps because, in multiple-case studies, documentation of treatment characteristics would be too time-consuming of a task. But the single-case approach can not only document the effects of the treatment but also the reasons for those effects. In fact, a single-case approach forces the evaluator to analyze the functional relationship between the treatment and the subject.

Example. Suppose an evaluator is assessing the effectiveness of computer assisted instruction (CAI) on only six students. He chooses to disaggregate his sample, and finds that the CAI was highly successful for five of the six students he had tried CAI with. Now, the five successful cases may represent a range of academic abilities, may have varied in their experience with automated devices, and may have come from different classrooms. How, then, does the sixth student differ? This example could have any number of endings. The sixth student may have had a slightly lower ability than the lowest of the other five, leading the evaluator to hypothesize that a minimal level of competence is necessary to function well with the machine. Or he may have had a previously undiscovered visual impairment which made words on a lighted screen difficult for him to read. Conversely, the evaluator might have discovered that each of the five "success" stories are due to unique features of these five students. They may have had a variety of emotional problems or family crises which were easier to escape from during computerized instruction than during classroom instruction. Any of these findings provide the evaluator with further clues as to how the treatment functions with regards to recipients. Our con-

cern in this section, then is to determine what the criteria are that allow generalization from a functional analysis of the treatment.

Just as the sample attributes are defined with reliance on prior knowledge, so are the treatment attributes. After all, the treatment was not developed in a vacuum. It was developed to meet a pre-conceived need for improvement of something, and it was developed on a hypothesis that for some (specified) reason this particular treatment would in fact accomplish the needed improvement. That is, the treatment was expected to influence cases in some particular way. The assumptions and goals of the treatment, then, provide a starting point for a functional analysis of its real influence. There are three criteria relating to the treatment attributes that can be used to increase either the strength or the range of generalization about its influence.

1. Wide Range of treatment attributes across replications. It is likely that administration of the treatment was varied slightly (or even greatly) across cases. That is, each case received a different variation of the treatment. The nature and extent of these variations may provide evidence of the treatment's utility across contexts, but more importantly they may allow the evaluator to identify those attributes of the treatment that appear to be functionally contributing to the outcomes. That is, they may allow the evaluator to separate the relevant from the irrevelant attributes of the treatment, so that conclusions may be based on more fundamental characteristics, rather than surface characteristics of the treatment. Now, to the extent that surface characteristics are defined, the treatment is less malleable and therefore less likely applicable in many different contexts. To

the extent that treatment variations allow more fundamental attributes to be identified, then, they also allow conclusions about effects to be stated in terms of the fundamental basis of the treatment and hence increase the range of generalization.

2. Common patterns of treatment outcomes across sample cases. The extent to which the treatment yields a similar set of outcomes or non-outcomes across the different cases lend credence to the hypothesis that the treatment did in fact have a predictable influence on the cases, even if the influence was not what was hypothesized. This criterion is further strengthened, of course, if the consistent patterns are observed in a sample of cases with varying attributes. It should be clear by now that the value of these criteria is not in their individual merits but rather in their collective use. This criterion is no exception. A replicated pattern of outcomes may do more for one's causal inferences than for one's generalization inferences, if it is not accompanied by some of these other criteria.

3. Common treatment functions across cases. This final criterion for the treatment, its functional influence, refers to the reason for the treatment effect. This criterion is our strongest criterion, but it too relies on other criteria. The advantage of the single-case methodology is that it forces the evaluator to look at the functional relationships between the treatment and the subject(s). Since the evaluator must assess the treatments influence on each case individually, he is more likely to discover the fundamental characteristic of the treatment and the extent to which the nature of treatment influence was related to sample or context attributes. Ultimately,

it would not be a treatment effect which was generalized -- it would be a relationship between treatment, context, and recipient that was generalized. And since the treatment may be re-cast in its more fundamental form, the conclusion may be applicable to a wider range of contexts or subjects.

Example. Suppose an evaluator is asked to access the effects of a "trial" trip to the zoo on children's attitudes toward school. The evaluator considers this to be a single-case study with the classroom as the single unit receiving tne treatment. He accompanies the teacher and her children to the zoo, observes children and interviews them about their attitudes. Every child he speaks to is delighted by the trip, but different children are pleased for different reasons. Bobby, for example, liked the trip because it was an escape from school, which he hated. Mike, on the other hand, liked the trip because it helped him understand his recent science lessons. Mike was anxious to return to school and continue his science lessons with the aid of his experience. The evaluator's functional analysis of the treatment allowed him not only to determine the effects of the treatment on attitudes, but also to determine the reasons for the effect. Further analysis of his data lead him to conclude that:

(a) the classroom was not the appropriate unit of analysis for two reasons. First, because the treatment was not functioning in the same way across children within the classroom, and second, because it was administratively possible to provide the treatment to individual children

rather than to classroom units.

(b) The hypothesized outcome of the treatment, i.e., to increase children's positive attitudes toward school, only occurred on a subset of children--those children which had never visited a zoo before. That is, it was the uniqueness of the experience that influenced the children.

The evaluator chose to generalize his findings, not only to all children who had never visited a zoo before, but also to other kinds of treatments, and concluded that:

(c) Excursions to places previously unfamiliar to the children, but which related to current curricular content, would probably produce similar attitudinal change.

Functional analysis allows the evaluator to redefine the treatment in a more fundamental way, and to describe the treatment effect in relationship to both context and recipient.

Generalizing from Single-Case

The foregoing has shown the advantages of replicated single-case designs. Disaggregation allows a functional analysis of the treatment, while replication allows an analysis of variability and commonality of sample attributes. Clearly a study of a single-case with no replications limits the strength of generalization arguments considerably. It does not preclude functional analysis, however, nor does it preclude a description of the relevant common and unique attributes of

he case. But, in fact, the range of generalization simply cannot be known to the evaluator.

That the range cannot be known, however, does not mean that a range does not exist. I mentioned earlier that the range of generalization was necessarily a matter of judgment, and I would now like to suggest that for studies of single cases, the judgment should not be made by the evaluator. Instead, it should be made by those individuals who wish to apply the evaluation findings to their own situations. That is, the evaluator should produce and share the information, but the receivers of the information must determine whether it applies to their own situation. Since the evaluator cannot know who his receivers are, he must, of course be quite specific both in his description of the attributes of his case and in his description of the way in which the treatment influenced this case.

Researchers and evaluators are not accustomed to the notion of leaving generalization up to the practitioner, but it is not an uncommon occurrence in other fields. In fact, some fields even specify criteria for generalization. Two, in particular, are worth reviewing: Legal generalizations and clinical generalizations.

Legal Precedent

The term "case law" refers to that portion of the law that is built up from specific cases, rather than from statutes. These specific cases are resolved on the basis of statutes, but the interpretations of statutes that are made in each case set precedents for future cases. If decisions are described in terms of general ideas, these ideas may become principles and take on a life of their own (Cardoza,

1921). Though these decisions may be stated with the intention that they be generalized, it is the later court which must decide whether in fact a particular decision should be applied to its own case. Thus, it is the receiver of the information who determines its generalization to a new situation. For that reason, the rules by which these judgments of generalization are reached might be useful to the educational decision-maker who needs to judge the generalizability of a single-case study to his own situation.

How are those decisions made? The process is one of search and comparison (Cardoza, 1921), in which the attributes of the current case are compared with the attributes of a variety of other cases. That case which is most analogous, that is, which has the most similar attributes is selected as the most revelant precedent. The process, then, is similar to that which we have already described for generalizing, and its validity hinges on the extent to which the attributes compared are relevant. Legal tradition focuses primarily on these attributes:

(a) the material facts of the earlier case (are they similar to the current case?)

(b) Appropriateness of the decision made in the case (would a similar decision here be consistent with one's sense of justice and social welfare?)

(c) the reason for the decision (the statute used to justify the decision, and/or the justice's reformulation of the statute)

(d) the level of generality with which the decision was formulated (the extent to which an intent to generalize seems apparent)

Let's consider how each of these might be applied by an educational decision-maker.

(a) Material Facts. In a court of law, material facts might consist of motivations, actions or transactions. In education, material facts may range from attributes of the recipients of treatments, to the treatment, to the context in which the study was done. They would also include the reasons for trying the treatment, i.e., the anticipated outcomes. The judicial process recognizes that not all material facts are relevant (just as we have also noticed) and provides a further stipulation that those facts which were relied upon in making the earlier decision are the most relevant facts. Suppose, for example, that an educational decision-maker is interested in improving children's attitudes toward school and he reviews the case study previously described. The relevant material facts are those used as a basis for the decision -- "children who have never visited a zoo" and "congruence between subject matter of the excursion and subject matter in the classroom".

(b) Appropriateness of the Decision. In a legal matter, as in an educational matter, appropriateness is determined by one's own value system. In law, the criterion is one's sense of justice or social welfare. In education, the criterion might be the same, or it could be one's sense of the purpose or goals of education. It might

seem odd to think that enough disparity could exist between the former case and the current one that the earlier decision could be completely inappropriate now, but this criterion is an important one for it recognizes the fluidity of our social fabric and the need to change precedents (and educational treatments) as social values change.

(c) Reason for the Decision. A court decision will necessarily be based on a statute or a re-formulation of a statute and may be justified by demonstrating the relationship between the material facts in the case and the intent of the statute. This reasoning is then used by the later court to determine whether the earlier case provides a relevant precedent. In education, laws are not available to base case-study conclusions on, but a functional analysis of the treatment may provide a similar criterion in the sense that it, too, provides a justification for reaching a decision as to the effects of the treatment. Suppose, for example, an educational decision-maker is contemplating the use of computer-assisted instruction and reviews the replicated single-case study described earlier. Suppose, further, that the earlier decision had been that CAI was not useful because only those children who had emotional problems benefited from it. If our current decision-maker happens to be in a school for delinquent boys, he may determine that the reason for rejecting CAI earlier did not hold for his current situation. Since many of his students had emotional problems, CAI may be valuable in his situation.

(d) Generality of the Decision. A legal example of generality has been shown by Gottlieb (1968) in a negligence case. A bottler was considered negligent because a consumer had found a dead snail in a

bottle of beer. The judge could have argued negligence for that specific reason, or because a dead creature was in a sealed bottle, or because a foreign object was in a container. The latter is clearly a more general statement than either of the others. Similarly, our evaluator of the effects of the excursion to the zoo chose to make a general decision about the relationship between the curriculum, the treatment, and the children's past experience, rather than simply saying that trips to the zoo were beneficial when animals were being studied and children had never been to a zoo.

These four criteria for generalizing from cases are clearly judgmental -- they are designed to provide guidance to users of information, not to those who generate it. In law, these two groups are synonymous, for a given judge may be a precedent-setter in one case and a precedent-user in another. The fact that he is sometimes a user makes him conscious of the potential implications of his decisions when he is a creator. Educational evaluators may be at a disadvantage if they operate as precedent-setters and remain segregated from the population of decision-makers who might wish to generalize their decision to other contexts.

Clinical treatments

Medical and clinical psychological professions serve individual clients and much of their clinical knowledge develops from the accumulation of findings about treatment effects on individuals. Individual cases may be studied to learn more about etiology, more about particular treatment effects, or because the cases are unique and need to be carefully studied to determine whether in fact the patient has a new,

as yet unrecognized ailment. To the extent that the purpose of the study is to facilitate classification, no inference is necessarily involved. We will confine ourselves here to those cases where an inference of generalizability is involved. Like generalizations in law, clinical generalizations are the responsibility of the receiver of information rather than the original generator of the information and so that evaluator must be careful to provide sufficient information to make such generalizations possible. Small and Krause (1972) have identified three important criteria for inclusion in a clinical report.

(a) longitudinal information (both extensive case history and extensive follow-up after treatment)

(b) Multi-disciplinary assessment of patients (Representation of a variety of specialties and perspectives)

(c) Precision in description (rather than imprecise or vague terminology)

Let us consider the application of these criteria to educational case studies.

(a) Longitudinal information. In clinical research, longitudinal data may provide information regarding the past (onset of symptoms, frequency or nature of other illnesses, etc.) or the future (recidivism, side effects, etc.) These findings place the clients' illness and treatment in a context of his general health and well-being. In education, longitudinal data provide a similar service in the sense that the function of the treatment vis-a-vis the subject is understood

relative to the history and development of that subject or context. Such information helps the user of the case study to estimate the similarity between the evaluator's case and his current case.

(b) Multidisciplinary Assessment. In medical inquiry, the use of a variety of specialists insures that all aspects of the clients' health are considered and thereby diminishes the possibility of unwitting mis-diagnosis or of confounding of different health problems. In education, similar advantages may be obtained by use of educational psychologists, sociologists, and developmental psychologists, each of whom may bring their own perspectives and interpretations to bear on the case. Use of several disciplines increases the breadth of information generated regarding the case and, hence, assists the user to know the full range of circumstances surrounding the case.

(c) Precision of description. Rather than simply describing a client's ailment by its categorical label, full description of its nature and degree communicate more precisely the client's situation. An application of this principle to education might mean that, for example, a child's reading ability would be described for oral versus silent reading, reading of different kinds of material, interest in reading, and so forth, rather than described simply by a grade-equivalent score.

In both legal and clinical fields, then, we see that generalizations are frequently necessary from single cases, but it is also clear that these generalizations are done by the user of the case data rather than by the person who originated the case data. And the generalization is not from a case to a population but rather from a case to another case.

Because the generalization is from one case to another, the user must rely on as much information as possible to determine the ways in which the two cases are analogous.

This approach to generalization seems appropriate to the field of education, for education is a highly decentralized industry with great variability among different practicing agencies. Though the evaluator may not care for the ambiguity of this situation, and may feel obligated to define his range of generalization precisely, he must not forget that regardless of his care, users will ultimately make their own decisions as to whether his findings are applicable in their situations. In fact, the user will probably, like the judge in court, study an array of available examples and pick the one which most closely approximates his own situation. To the extent that this is so, single-case studies will prove more valuable to educational decision-makers than group studies since these studies may not allow generalization to individual cases.

Summary

The importance of generating some set of rules for non-statistical inferences should not be underestimated. Part of the value of grouped studies is that well-defined statistical rules have been developed to aid our inferences. The clarity and specificity of these rules makes it relatively easy for different people to agree on the generalizability of a study. It is no trivial matter to embark on a new course for which the rules of inference have not been established. For that reason, I have tried to offer a starting point for the development of an alternate set of rules. Surely these can be improved and other rules can be added, and just as

surely, we are a long way from approaching the sophistication that our statistical rules have realized.

An important topic which I have not addressed is the relative strength of these different rules, and I have avoided that issue intentionally for I think their relative utility may depend on the subject being investigated and the available knowledge and theory about that subject.

The rules suggested here may appear to require a larger dose of subjective judgment than statistical rules do, but this may not be the case in practice. For in conducting a multiple-case study, judgments are made as to how to sample, how to design the study, and what type of statistical analyses to employ. Furthermore, since samples are rarely drawn randomly from the population to which inferences are desired, judgments are used for the second span of Cornfield and Tukey's (1956) two-span bridge of inference.

Judgments must be similarly employed in single-case studies, but they need not be whimsical. Though educational program evaluation as a field has existed for little more than a decade, a considerable body of theory and knowledge exists that can provide a basis for these judgments.

References

Campbell, D. T. and Stanley, J. Experimental and quasi-experimental designs for research on teaching. New York: McGraw-Hill, 1963.

Cardoza, B. N. The nature of the judicial process. Princeton, N. J.: Yale University Press, 1921.

Cornfield, J. and Tukey, J. W., Average values of mean squares in factorials. Annals of Mathematical Statistics, 1956, 27, 907-949.

Dukes, W.F. N = 1. Psychological Bulletin, 1965, 64, 74-79.

Edgar, E. and Billingsley, F. Believability when N = 1. The Psychological Record, 1974, 24, 147-160.

Edgington, E.S. Statistical inference from N = 1 experiments. The Journal of Psychology, 1967, 65, 195-199.

Feinberg, S.E. The collection and analysis of ethnographic data in educational research. Anthropology and Education Quarterly, 1977, 8 (2), 50-57.

Gottlieb, G. The logic of choice. New York: McGraw-Hill, 1963.

Herson, M. and Barlow, D. Single-case experimental designs. New York: Pergamon Press, 1976.

McNemar, Q. Sampling in psychological research. Psychological Bulletin, 1940, 37, 331-365.

Rist, R.C. On the relations among educational research paradigms: From disdain to detente. Anthropology and Education Quarterly, 1977, 8 (2), 42-49.

Small, W.P. and Krause, U. An introduction to clinical research. London: Churchill Livingston, 1972.

Stake, R.E. The case study method in social inquiry. Educational Researcher, 1978, 7 (2), 5-8.